PREDICTION OF EYE, HAIR AND SKIN COLOR IN ADMIXED POPULATIONS OF LATIN AMERICA

Sagnik Palmal[1*], Kaustubh Adhikari[2,3*], Javier Mendoza-Revilla[4,5], Macarena Fuentes-Guajardo[6], Caio C. Silva de Cerqueira[7], Juan Camilo Chacón-Duque[8], Anood Sohail[9], Malena Hurtado[4], Valeria Villegas[4], Vanessa Granja[4], Claudia Jaramillo[9], William Arias[10], Rodrigo Barquera Lozano[11,12], Paola Everardo-Martínez[11], Jorge Gómez-Valdés[11], Hugo Villamil-Ramírez[13], Tábita Hünemeier[14], Virginia Ramallo[15,16], Rolando Gonzalez-José[16], Lavinia Schüler-Faccini[15], Maria-Cátira Bortolini[15], Victor Acuña-Alonzo[11], Samuel Canizales-Quinteros[13], Carla Gallo[4], Giovanni Poletti[4], Gabriel Bedoya[10], Francisco Rothhammer[17,18], David Balding[3,19], Pierre Faux[1#], Andrés Ruiz-Linares[1,3,20#]

[1] UMR 7268 ADES, CNRS, Aix-Marseille Université, EFS, Faculté de Médecine Timone, Marseille, 13005, France.

[2] School of Mathematics and Statistics, Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes, MK7 6AA, United Kingdom.

[3] Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, WC1E 6BT, UK

[4] Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, 31, Perú.

[5] Unit of Human Evolutionary Genetics, Institut Pasteur, Paris, 75015, France.

[6] Departamento de Tecnología Médica, Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, 1000000, Chile.

[7] Scientific Police of São Paulo State, Ourinhos-SP, 19900-109, Brazil.

[8] Division of Vertebrates and Anthropology, Department of Earth Sciences, Natural History Museum, London, SW7 5BD, UK.

[9] Department of Biotechnology, Kinnaird College for Women, 93 - Jail Road, Lahore, 54000, Pakistan.

[10] GENMOL (Genética Molecular), Universidad de Antioquia, Medellín, 5001000, Colombia.

[11] National Institute of Anthropology and History, Mexico City, 6600, Mexico.

[12] Department of Archaeogenetics, Max Planck Institute for the Science of Human History (MPI-SHH), Jena, 07745, Germany.

[13] Unidad de Genomica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM-Instituto Nacional de Medicina Genómica, Mexico City, 4510, Mexico.

[14] Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508-090, Brazil.

[15] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, 90040-060, Brasil.

[16] Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico, CONICET, Puerto Madryn, U9129ACD, Argentina.

[17] Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Arica 1000000, Chile.

[18] Programa de Genetica Humana, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Arica 1000000, Chile.

[19] Melbourne Integrative Genomics, Schools of BioSciences and Mathematics & Statistics, University of Melbourne, Melbourne, VIC 3010, Australia.

[20] Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Yangpu District, Shanghai, China.

*Contributed equally to this work. #Co-supervised this work
Corresponding authors: andresruiz@fudan.edu.cn; pierre.faux@univ-amu.fr

Abstract

We report an evaluation of prediction accuracy for eye, hair and skin pigmentation based on genomic and phenotypic data for over 6,500 admixed Latin Americans (the CANDELA dataset). We examined the impact on prediction accuracy of three main factors: (i) The methods of prediction, including classical statistical methods and machine learning approaches, (ii) The inclusion of non-genetic predictors, continental genetic ancestry and pigmentation SNPs in the prediction models, and (iii) Compared two sets of pigmentation SNPs: the commonly-used HIrisPlex-S set (developed in Europeans) and novel SNP sets we defined here based on genome-wide association results in the CANDELA sample. We find that Random Forest or regression are globally the best performing methods. Although continental genetic ancestry has substantial power for prediction of pigmentation in Latin Americans, the inclusion of pigmentation SNPs increases prediction accuracy considerably, particularly for skin color. For hair and eye color, HIrisPlex-S has a similar performance to the CANDELA-specific prediction SNP sets. However, for skin pigmentation the performance of HIrisPlex-S is markedly lower than the SNP set defined here, including predictions in an independent dataset of Native American data. These results reflect the relatively high variation in hair and eye color among Europeans for whom HIrisPlex-S was developed, whereas their variation in skin pigmentation is comparatively lower. Furthermore, we show that the dataset used in the training of prediction models strongly impacts on the portability of these models across Europeans and Native Americans.

Keywords: DNA phenotyping, eye-color, hair-color, skin-color, pigmentation prediction, admixture, Latin Americans

Introduction

There is growing interest in the use of genetic data for the prediction of physical appearance, particularly in forensic, historical and paleo-anthropological studies[1–3]. Strong impetus for these studies has been provided by Genome Wide Association Studies (GWAS) of traits such as pigmentation, which have enabled the identification of single nucleotide polymorphisms (SNPs) associated with eye, hair and skin color[4–6]. Early prediction studies evaluated the accuracy of predicting eye color from SNP data[7–10] and these analyses have been more recently extended to hair [11,12], and skin pigmentation[13]. As a result, sets of SNPs have now been proposed for the simultaneous prediction of eye, hair and skin color[14,15]. However, so far, these studies have had a strong European bias[7,10,11,14,16]. The pigmentation SNPs for prediction have been selected from GWAS performed mainly in Northern Europeans[4,5,17] and prediction models also optimized mostly in European datasets[17–21]. Furthermore, evaluation of the accuracy of these prediction tools has also been performed essentially in European-derived population samples[18–20,22].

Among major continental populations, Europeans show the largest variation in eye and hair color, whereas the variation in skin color in Europeans is relatively lower compared to other continental regions such as Africa or South Asia[23–28]. The other continental regions have narrower variation in eye and hair color, mostly within the brown-black range; although fine-scale quantitative measurement still allows us to discern this variability and identify genetic variants unique to those populations[27,29–31]. Interestingly, even though the underlying biological mechanisms are broadly similar for these different types of pigmentations, their genetic architecture is substantially different, as eye and hair pigmentation are chiefly controlled by a handful of large-effect variants, whereas skin pigmentation is observed to be more polygenic and the impact of any single variant is smaller[6,16,24–27]. Consequently, studies assessing the portability of pigmentation

prediction systems derived primarily on European cohorts have reported good portability for eye and hair phenotypes in other non-European continental populations, where most samples were correctly predicted to be in the brown-black category, whereas portability for skin color in other continental populations was poorer[23,32,33]; for example, all light-skinned Japanese individuals were incorrectly predicted to be of dark skin color[23]. The convergent evolution of lighter skin color in different regions of Eurasia, as reported previously in the CANDELA cohort[31] and elsewhere[26,34,35], is the likely cause for this discrepancy, and highlights the importance of using diverse cohorts for building skin pigmentation prediction models.

Latin Americans represent one of the largest recently admixed populations world-wide. The history of Latin America has involved extensive admixture, particularly between Native Americans, Europeans and sub-Saharan Africans (with respective estimates of ancestry share: 38%, 52% and 6%[36]). Consistent with its partly European ancestry, a recent GWAS for pigmentation traits in Latin Americans in the CANDELA cohort detected phenotypic effects for a number of loci previously identified in Europeans[31]. In addition to these, novel pigmentation SNPs with genome-wide significant association were also identified in that study. These included SNPs polymorphic only in East Asians and Native Americans, consistent with the independent evolution of skin pigmentation in West and East Eurasia[26,34,35]. The admixed ancestry of Latin America and the finding in the region of pigmentation variants not present in Europeans emphasizes the need to evaluate the accuracy of tools currently available for prediction of pigmentation traits in this population.

Here we aimed to evaluate the accuracy of prediction of pigmentation traits in Latin Americans in the large Latin American dataset (the CANDELA cohort) recently used in GWASs of physical appearance, including for hair, eye and skin pigmentation[31,36–41]. We focused on comparing the accuracy of prediction methods, the predictors included in models,

and the SNP sets used for prediction. We find that Random Forest and regression are generally the method with best performance, depending on the predicted trait. We also find that the inclusion of pigmentation SNPs increases the accuracy of prediction models substantially over that obtained by the inclusion of other predictors. Finally, the HIrisPlex-S has a similar performance to a CANDELA-specific set of prediction SNPs except for skin pigmentation.

Materials and Methods

Study sample: phenotypes, genetic data and covariates

We analyzed data previously studied by the CANDELA consortium for GWAS of pigmentation traits[31,36–39]. The consortium gathered genetic and phenotypic data from over 6,500 individuals recruited in five Latin American countries: Mexico (N=~1,200), Colombia (N=~1,700), Peru(N=~1,230), Chile (N=~1,730) and Brazil (N=~630).

Pigmentation traits evaluated directly on the research subjects consists of (A) hair pigmentation (recorded in four categories: 1-red/reddish, 2-blond, 3-dark blond/light brown or 4-brown/black. However, due to their very low frequency (<0.6%), individuals in the 'red/reddish' category were not included here), (B) eye color, recorded as five ordered categories: 1-blue/gray, 2-honey, 3-green, 4-light brown, 5-dark brown/black. For increasing consistency with previous publications[10,13,42–44], here we recoded these data into just three categories: 1-Blue/Gray, 2-Intermediate (honey or green) and 3-Brown/Black (light brown or dark brown/black) and (C) a quantitative measure of skin pigmentation from an area unexposed to sunlight (the Melanin Index MI, obtained by reflectometry). We also had available additional measures of iris pigmentation, extracted from digital photographs, using the HCL color space (Hue, Chroma and Luminance). Hue being an angle (recorded in arc degree), we linearized this trait with cosine and shifted the angle by 15° in order to maximize

the number of samples in the range [0,180°]; hence the trait considered is cos(Hue+15). The frequency distribution for these traits in the CANDELA dataset is shown in Supplementary Figure S1.

For comparison with other studies[13,42], we converted the skin MI into a three-level categorical trait. For this, we selected CANDELA individuals with 100% European ancestry (N=70) and examined their facial photographs in order to label them as very pale or pale. Similarly, we selected CANDELA individuals with highest African ancestry (>39%; N=23) and examined their photographs in order to label them as dark or very dark. We plotted the distribution of the MI for individuals in the very pale, pale, dark and very dark categories and identified MI values of 33 and 47 as thresholds for categorical skin color: fair (MI <33; N= 2,506), intermediate (MI 33-47; N=3,840) and dark (MI>47; N=180) classified as dark (see Supplementary figure S2). These thresholds are in line with values obtained in another study of admixed Brazilians[45].

The genetic data consisted of ~9 million genotypes, ~700k of which were obtained experimentally by genotyping Illumina's Omni Express chip, the remainder obtained by imputation as described in Adhikari et al[31]. We applied several filters to the CANDELA dataset prior to the trait prediction analyses. Firstly, we retained only individuals aged 18 to 45. Secondly, we removed 8 pairs of individuals whose pairwise probability of IBD was estimated close to 1, to discard potential sample mix-ups (hence, 16 individuals removed), and individuals whose estimated African ancestry was more than European and native ancestry estimations (23 individuals), as those were considered as genetic outliers and thus excluded. Finally, we excluded all individuals with missing data on any of the covariates (age, sex, BMI). Note that BMI was considered as a covariate since we found it significantly correlated to some pigmentation phenotypes; that correlation is most likely a confounding effect of continental genetic ancestry. The final sample size used in the analyses was: 6,495

for hair color, 6,526 for MI, 6,529 for categorical eye color and 5,738 for quantitative eye color traits – Hue, Chroma and Luminance. These three eye color phenotypes constitute the bicone color space model – HCL scale for human perception of eye colors (previously explained by Adhikari et al[31]).

Pigmentation SNP sets used for prediction

We used two sets of SNPs for the prediction analyses of pigmentation traits.

Firstly, as a benchmark we used HIrisplex-S, a SNP set that has been developing over the years for the prediction of eye, hair and skin pigmentation mainly in Europeans[7,11,14] HIrisplex-S currently includes 41 SNPs, 22 of which were directly genotyped in the CANDELA samples (genotypes for the remaining 19 SNPs being imputed). In our analyses, we only retained SNPs with MAF >= 1% in the CANDELA data, reducing the HIrisplex-S set used here to 34 SNPs (see Supplementary table S1).

Secondly, we devised "CANDELA" (CAN) SNP sets for prediction of each pigmentation trait (E-eye; H-hair; S-skin) based on results from a GWAS conducted in the CANDELA sample[31]. To pre-select SNPs for each trait, we used the following protocol: (1) selection of all SNPs with GWAS association p-values <1E-5, (2) grouping (clumping) of SNPs in high LD and (3) for each SNP group, selection of the SNP with highest predictive power (it is well-known that the most significant marker is not necessarily the best predictive marker[46]). In this way we pre-selected 1,471 SNPs relevant for skin pigmentation prediction, 207 for hair pigmentation prediction and 701 for eye pigmentation prediction. These SNPs were then ranked based on conditional predictive power for each trait, in order to search for the set of predictive SNPs with the least SNPs included. Details of the approach used are provided in Supplementary method S1, and the resulting CAN-E, CAN-S and CAN-H sets described in the Results section.

Prediction methods and models evaluated

A broad array of statistical methods have been employed in the literature to predict pigmentation traits, such as (multiple) linear[17,47] or (multinomial) logistic regression[10,11], decision trees[17,48], neural networks[17,49], and naïve Bayes classifiers[33,42,50]. Each method has its advantages and disadvantages, and are better suited for certain types of traits, e.g. linear regression for quantitative traits[17,47] and logistic regression for categorical traits[10,11].

The overall strategy we used for performing pigmentation prediction in the CANDELA dataset is shown in Figure 1. Linear Regression (LR) or Multinomial Logistic Regression (MLR) were used as the reference methods for quantitative or categorical traits, respectively. These two methods were used to evaluate three prediction models, incorporating an increasing number of predictors (Fig. 1A):

1- Using only non-genetic covariates as predictors:

$$y \sim Age + Sex + BMI \qquad \text{(equation 1)}$$

2- Incorporating genetic ancestry to model 1:

$$y \sim Age + Sex + BMI + EUR_{ancestry} + AFR_{ancestry} \qquad \text{(equation 2)}$$

Here we included as predictors the estimates of European and African ancestry (obtained by unsupervised admixture estimation on genome-wide data). Native American ancestry was omitted so as to avoid collinearity (since the three continental ancestries sum to 1).

3- Incorporating pigmentation SNPs to model 2:

$$y \sim Age + Sex + BMI + EUR_{ancestry} + AFR_{ancestry} + \sum_{j \in SNPsets} SNP_j, \text{(equation 3)}$$

where SNPsets refers to SNPs included either in the HIrisPlex-S or CAN sets defined above.

For the third (full) model, in addition to regression, the following statistical and machine-learning methods were used, in order to evaluate their relative performance for prediction: Random Forest (RF), Extreme Gradient Boosting (XGB), Artificial Neural Network (ANN), Ordinal regression (OR) and Stepwise regression (SR). We provide more details on their implementation in Supplementary method S2.

Figure 1. Study overview. (A) Models tested, predictors used and prediction methods: multinomial regression (MLR), linear regression (LR), ordinal regression (OR), stepwise regression (SR), random forests (RF), extreme gradient boosting (XGB) and artificial neural network (ANN). **(B)** 10-fold cross-validation: the full data is randomly split into 10 equally-sized data sub-groups. For each of the 10 sub-groups, the estimation of model parameters **(C)** or optimization of model hyperparameters **(D)** was performed on a pool of the nine remaining sub-groups.
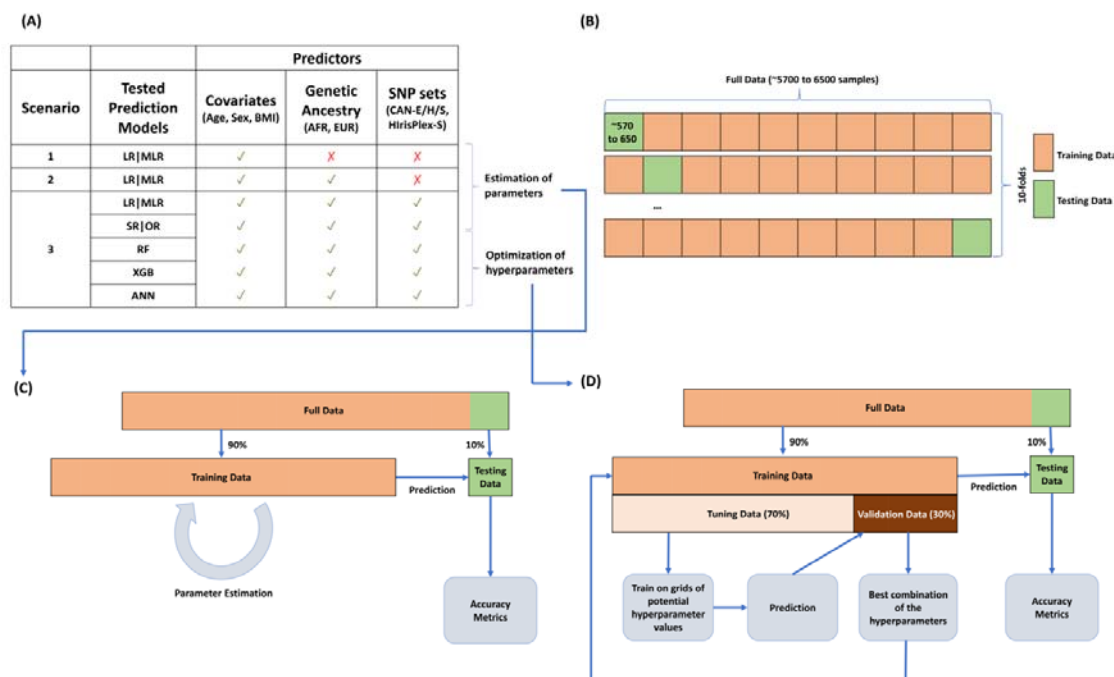
Evaluation of prediction accuracy

To measure prediction accuracy in quantitative traits, we used the coefficient of determination (R^2, the proportion of phenotype variance that is explained by the model, measured as $1 - SS_{res}/SS_{tot}$, where SS_{res} and SS_{tot} respectively stand for the residual and

total sum of squares). For categorical traits, we used a metric denoted "accuracy" that is the proportion of correctly classified individuals in confusion matrices. For these traits, we also computed the Area Under the ROC Curve (AUC – ranging from 0.5 to 1) as well as the expected accuracies from two benchmark strategies (either using the categories' frequency – PropStrat, or always guessing the most frequent category – maxP; see Supplementary Method S3) for comparison.

Accuracy of prediction was evaluated using 10-fold cross validation (10-fold CV) (Fig. 1B). For this, the full dataset is split into 10 approximately equal subsets based on a stratified sampling on the trait (so that trait distribution is similar across subsets). Each of the $1/10^{th}$ subsets is used as test data for evaluating prediction accuracy of methods trained using the other $9/10^{th}$ of the data. For regression methods (MLR, LR, SR, OR), coefficient parameters are estimated in the training data (Fig. 1C). For Machine Learning models, the training data is further split into a tuning data (70% of the training data) and validation data (30% of the training data). The parameter space for these methods are tuned creating a grid of all possible combinations of the hyperparameters and the particular combination producing best result on the validation data is selected as the set of optimal combination for the whole training data (see Supplementary Method S2). All possible permutations of binding nine folds for training produces 10 different train-test data combinations and the hyperparameters are tuned based on that subsequently gives rise to 10 prediction accuracy results (Fig. 1D). These measures of goodness of fit are used in a boxplot, or the average of them is used as a single prediction accuracy metric of the method. This helps us in avoiding inflation in the results and the predictions are more robust to small changes in the data.

Prediction of MI in Native American individuals of unknown phenotype

Genotype and geo-localization data for 104 Native American individuals (with an estimated >99% Native American ancestry) from 16 Native American populations were available from a previous study[40]. We analysed genotypes for these 'pure' Native American individuals, predicted their MI, and regressed these predicted values on the amount of solar radiation at the site of population sampling. We trained two RF models (one with CAN-S and one with HIrisPlex-S) using 550 CANDELA individuals with >= 80% native ancestry and using sex as the only covariate. Solar radiation levels were defined as insolation incident on a horizontal surface (in $kWh/m^2/day$) as reported in the NASA Surface meteorology and Solar Energy (SSE) Web site (https://eosweb.larc.nasa.gov/sse/) (data previously used in[31]).

Results

Below we compare the performance of HIrisplex-S with SNP sets devised here for the prediction of eye (CAN-E), hair (CAN-H) and skin (CAN-S) from summary statistics of a pigmentation GWAS performed in the CANDELA sample (Materials and Methods and Supplementary method S1). The sets devised here consist of: 56 (CAN-E), 101 (CAN-H) and 120 (CAN-S) pigmentation-associated SNPs and are detailed in Supplementary table S2. Consistent with the genetic correlation of eye, hair and skin pigmentation, some SNPs are shared across the CAN-E/H/S sets, as well as with the HIrisPlex-S set. The overlap between these four SNP sets is shown in Figure 2.

Figure 2. Overlap between SNP sets used for prediction of pigmentation traits. CAN-E, CAN- S and CAN-H refer, respectively, to SNP sets designed here for the prediction of eye, skin and hair pigmentation, based on a GWAS performed in the CANDELA sample[31]. HIrisPlex-S is a SNP assay developed for simultaneous Eye, Hair and Skin color prediction[14]. Numbers refer to SNPs shared between SNP sets.
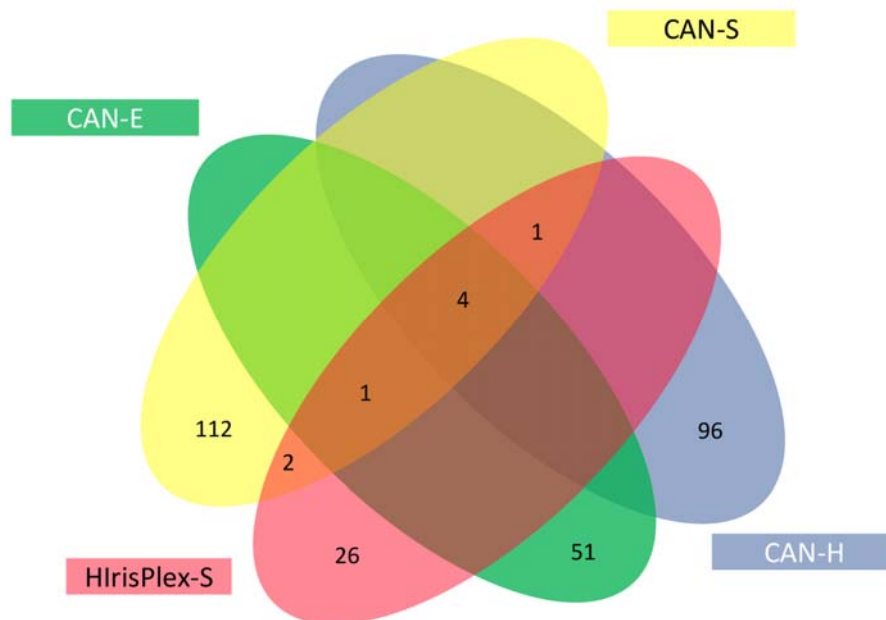

Prediction Accuracy in relation to models, methods and pigmentation SNP sets

Figure 3 presents the accuracy of prediction for various phenotypes of eye, hair and skin color. For categorical traits, the baseline model (i.e. including only non-genetic predictors: age, sex and BMI), reaches 84.9% and 81.7% accuracy (proportion of correctly classified individuals) for eye and hair color, respectively. That level of accuracy is actually also reached by always guessing the phenotype to be the most frequent category (maxP strategy, see Supplementary table S3 and Supplementary Method S2). This high accuracy obtained by a deterministic strategy probably relates to the highly skewed trait distribution in the CANDELA individuals: ~ 82% having black/dark brown hair and 85% having Brown/Black eyes. Alternately, randomly guessing the phenotypes based solely on the

frequency of the traits (PropStrat; cyan line in Figure 3) also yields good levels of accuracy

(~74% and ~69% for categorical eye and hair color, respectively).

Figure 3. Prediction accuracy in relation to models, methods and pigmentation SNP sets. For continuous traits (Hue (transformed), Luminance, Chroma and Melanin Index; top and middle panels) we used R^2 as measure of prediction accuracy. For categorical traits (Eye and Hair color; bottom panels) accuracy is the proportion of correctly classified individuals. Magenta and blue lines indicate the accuracy obtained when only non-genetic predictors or non-genetic + genomic ancestry are included in regression models, respectively. For categorical traits, the performance of a random guessing strategy (PropStrat) was also evaluated (cyan line). For these traits, the average accuracy of the deterministic maxP strategy is numerically the same as the accuracy obtained when only non-genetic predictors are used (magenta line), hence is not shown separately in this figure. For the full prediction model (non-genetic predictors + genetic ancestry + pigmentation SNPs) the performance of regression and four additional prediction methods was evaluated (bars are colored: green = LR/MLR; yellow = SR/OR; brown = RF; pink = XGB; purple = ANN). Detailed numerical values are given in Supplementary Table S3. The pigmentation SNP set incorporated in the prediction models is indicated at the bottom of the plots.
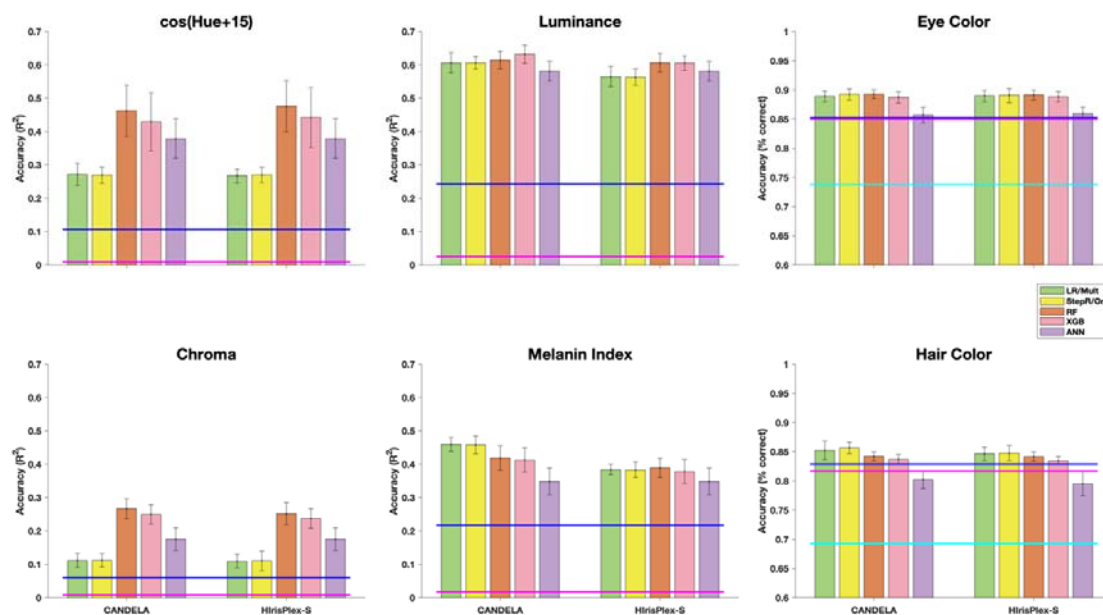
It is important to keep the maxP strategy in context when assessing prediction

performance for categorical traits, since it represents how skewed the trait distribution is – a

binary trait with a frequency distribution of 90% and 10% of the two categories will have

90% accuracy under the simplest maxP strategy (even though its sensitivity will be 0 for the

rare category; see Supplementary Method S2). Thus, a skew in the trait distribution causes an upward shift in accuracy of prediction methods, especially those methods which are biased to the most frequent category, making them appear better-performing than they actually are. The PropStrat strategy is comparatively less biased as it gives proportional weight to the rare category (hence non-zero sensitivity for this category), and thus has lower accuracy than the maxP strategy. It is therefore a better benchmark to compare the performance of other strategies for assessing their gain in accuracy. Conversely, a comparison of those strategies to the maxP benchmark better represents their relative change in incorrect classification rather than the relative gain in correct classification (i.e. gain in accuracy).

Although the accuracies of these basic strategies are already high due to our skewed trait distributions, adding genetic ancestry to the model has a further impact, especially for hair color: it decreases the proportion of error by ~6% (from 18.3% of error to 17.1% - detailed numbers in Supplementary table S3). Then the further addition of pigmentation SNPs has an even larger effect: the remaining proportion of errors decreases by another ~16% and ~27% respectively for hair and eye color. It is also noticeable that the gain in prediction brought by SNPs relatively to that brought by genetic ancestries is much larger for eye color (~14x) than for hair color (~3x).

For continuous traits, we observe a large increase in prediction accuracy (R^2) when genetic ancestry is incorporated in regression models, relative to the baseline (including only non-genetic predictors). Furthermore, when pigmentation SNP sets are incorporated in the regression model, accuracy usually more than doubles over that obtained with genetic ancestry plus non-genetic covariates (green bar versus blue line in Figure 3). When using this full prediction model, lowest LR prediction accuracy was observed for Chroma (R^2 ~0.12) and highest for Luminance (R^2 ~0.58), two quantitative estimates of eye color variation.

Comparing different prediction methods for the full model (i.e. incorporating all predictors) we do not observe large differences in performance, with the exception of a relatively lower accuracy of regression for Hue and Chroma. For those two traits, RF markedly outperforms regression methods, more than doubling the accuracy of LR in the case of Chroma. For the two other continuous traits (Luminance and Melanin Index), regression models are as effective as machine learning models (Figure 3). We also note that those rank similarly throughout categorical and continuous traits: RF is almost always better than the other tree-based model (extreme gradient boosting; XGB) and artificial neural networks (ANN) always underperform compared to tree-based models.

Regarding the SNP sets tested, we observe little difference in prediction accuracy across traits and methods although the number of SNPs is larger in the sets that we designed (56 to 120 SNPs) than in HIrisPlex-S (34 SNPs). Continuous skin pigmentation (Melanin Index) stands nonetheless as an exception: for that specific trait, CAN-S consistently outperforms HIrisPlex-S (particularly with regression methods). We have further assessed the impact of limiting the number of SNPs from the CAN sets to the top 10, 20 or 30 SNPs. A reduction in the number of SNPs had little effect for most traits, to the noticeable exception of Melanin Index for which considering the top 30 SNPs (i.e. 25% of the CAN-S SNP set) yields less than 90% of the prediction R^2 achieved with the whole set (see Supplementary table S7). Also, we ranked the SNPs by their variable importance (returned by the RF models) for each trait (see Supplementary table S4 and figure S3). Three SNPs, from three known pigmentation genes (SLC45A2, HERC2 and SLC24A5), are consistently among the top ones throughout traits and are common to HIrisPlex-S and the 3 CAN SNP sets.

**Prediction accuracy at varying levels of European/Native American ancestry**

Since a substantial fraction of individuals in the CANDELA sample have minimal African ancestry, we sought to evaluate prediction accuracy specifically for varying levels of European/Native American ancestry in the CANDELA sample. To this aim, we pooled individuals with negligible African ancestry in ~20% ancestry bins (so that the smallest pool size included at least 570 individuals) and examined prediction accuracy in the pools (see Supplementary table S5). Furthermore, for categorical traits, we ensured that at least two trait categories, each with >20 individuals, were observed in the pools, which led to withdraw the most Native-American pool. We assessed prediction accuracy using RF, a full model (i.e. equation 3 but without genetic ancestry as predictor) including only pigmentation SNPs having >1% MAF in the pool of individuals being tested.

For the categorical traits, there is a drop in prediction accuracy (from ~95% to ~70%) at increasing European ancestry (Figure 4). However, as European ancestry increases there is greater accuracy relative to random guessing based on trait frequency, probably reflecting the trait being less variable at higher Native American ancestry levels. For the quantitative eye pigmentation variables (particularly H and L, Figure 4), as the percentage of European ancestry increases there is a trend for an increase in trait variation (red line in Figure 4) and also in prediction R^2. For skin pigmentation (MI) we observe an opposite trend in trait variability in relation to ancestry, relative to hair/eye color: variation in MI decreases at increasing European ancestry. There is also a trend towards an increase in the performance of the CAN-S SNP set at decreasing European ancestry: in individuals with <20% European ancestry CAN-S has an accuracy that is nearly twice that observed for HIrisPlex-S. Although CAN-S tends to outperform HIrisPlex-S in most comparisons, it is only for MI that such a large difference in performance was observed. In summary, across all pigmentation traits we observe a gain in prediction accuracy for the Native American/European ancestry bins

showing greater phenotypic diversity, with CAN SNPs markedly outperforming HIrisPlex-S only for MI in individuals with low (<20%) European ancestry.

Figure 4. Prediction accuracy for individuals with varying Native American/European admixture. Prediction was assessed in individual bins varying ~20% in admixture (bottom axis; for eye and hair pigmentation <20 individuals with >80% Native Ancestry were available in each trait category, thus preventing estimation of prediction accuracy). Colored bars indicate accuracy obtained with Random Forest models using non-genetic + pigmentation SNP sets as predictors (R^2 being used as prediction measure for quantitative traits: Hue, Luminance, Chroma and Melanin index). Blue bars indicate the CAN-E/H/S SNP sets. Yellow bars indicate the HIrisPlex-S set. The standard deviation of the quantitative traits in each ancestry bin is indicated as a red line. For the categorical traits (Eye and Hair Color), accuracy (proportion of correctly classified individuals) is used as the metric for prediction measures. Accuracy obtained without genetic predictors using a guessing strategy is indicated with a horizontal blue line for Proportional Strategy (random guessing) and magenta line for maxP (deterministic guessing). Detailed numerical values are given in Supplementary Table S8 and S9.
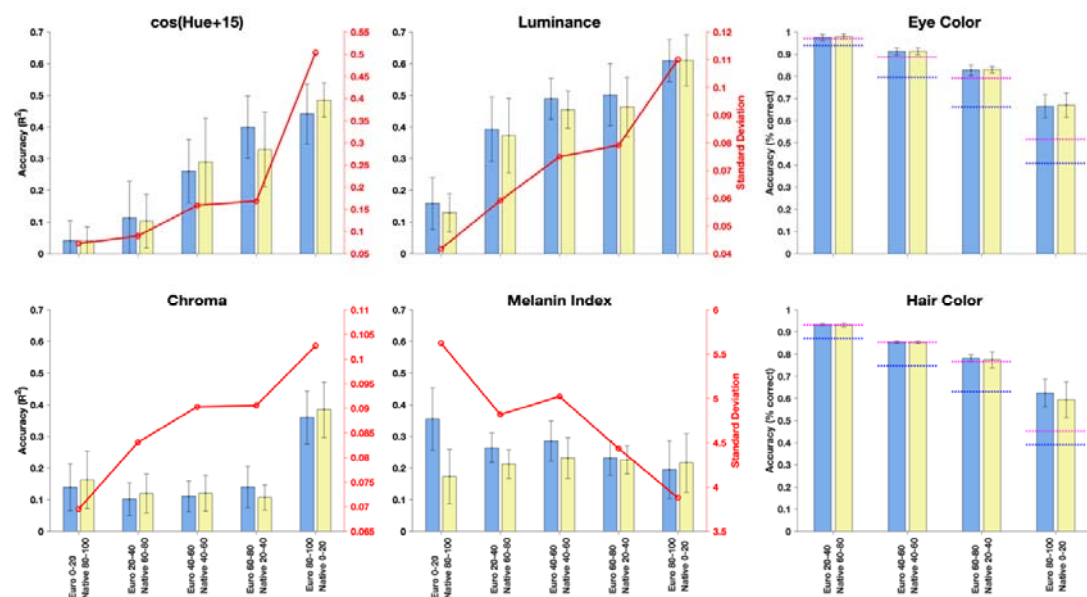
Prediction accuracy in CANDELA relative to other population samples

Several studies have examined accuracy of prediction of the HIrisPlexS SNP set for categorical pigmentation traits using an online tool (https://hirisplex.erasmusmc.nl, referred to here as HIrisPlexS-Online), which implements MLR prediction models trained in a reference dataset (consisting mostly of Europeans)[7,11,14]. Table 1 compares the published

estimates with those we obtained here with the CAN and HIrisPlex-S SNP sets using our implementation of MLR models trained in the CANDELA data (for skin color, we transformed MI into a 3-level categorical variable: fair, intermediate and dark; as described in Materials and Methods). For a more accurate comparison to HIrisPlexS-Online, these models were trained using SNPs only, without basic covariates or ancestry.

Prediction accuracy estimates for eye color have been reported for HIrisPlex-S-Online in a Latin American sample (99 individuals from Venezuela and Brazil) and in a European sample[10,14,15]. The light (blue/gray) and dark (Brown/Black) color categories have similar prediction accuracies across studies (~90-93%), except for the prediction of light eye color reported for HIrisPlex-S-Online in the Venezuelan/Brazilian sample, where accuracy is much lower (85%). The main difference in eye-color prediction across studies lies in the intermediate category. No intermediate eye colors were predicted by HIrisPlex-S-Online in the Venezuelan/Brazilian sample. Our predictions in the CANDELA sample have markedly higher accuracy than reported for HIrisPlex-S-Online in Europeans, both when the prediction model was trained in the CANDELA data or in the reference HIrisPlex-S data (respectively 89% and 85%, versus 73% in Europeans). Prediction accuracy, in the CANDELA sample, of the CAN-S and HIrisPlex-S SNP sets was identical.

Estimates for hair color prediction accuracy using HIrisPlex-S-Online have been reported for a European sample[44]. Prediction accuracy estimates obtained here for the CANDELA sample are higher than reported in Europeans for all hair colors, except in the case of intermediate hair-color (i.e. brown) predicted with HIrisPlex-S-Online. The highest hair-color prediction accuracy was consistently obtained with the CAN-H SNP set, although the difference relative to the HIrisPlex-S, with models trained in the CANDELA data, is small.

Concerning categorical skin color, we observe that predictions from HIrisPlex-S-Online have markedly lower accuracy in the CANDELA sample than reported for a world-wide sample. Model training in the CANDELA dataset increases prediction accuracy substantially both for HIrisPlex-S and CAN-S, although the accuracy values obtained are still below the published estimates for HIrisPlex-S. As for hair color, the CAN-S SNP set consistently outperforms HIrisPlex-S, but only marginally.

Portability of models for pigmentation prediction in individuals with high Native Ancestry

Considering the impact of training datasets in the performance of HIrisPlex-S (Table 1), we specifically examined the portability of RF models developed in two training datasets with extreme differences in ancestry (extracted from the CANDELA sample – see Supplementary figure S4): (i) a highly European training dataset (European ancestry >= 80% and Native American ancestry < 20%) and (ii) a highly Native training dataset (European ancestry < 20% and Native American ancestry >= 80%). We examined the performance of the resulting prediction models in a subset of the highly Native test dataset (Figure 5) in a cross-validation scheme. We observe that models developed in the highly Native training dataset have a better performance than those developed in the highly European training dataset for Chrome, Luminance and MI. The most striking difference in performance is seen for MI, where the model trained with highly Native data has a prediction accuracy ~6 times that of the model trained in highly European data (Figure 5). Hue is the only trait for which the model trained in the highly European dataset outperforms the model trained in the highly Native dataset, but prediction accuracy in this case is extremely low (<2%).

Figure 5. Portability of prediction models trained in highly European/Native American cohorts. For each continuous trait (cos(H+15), C, L and Melanin Index) we compare prediction accuracy on the same test data (a highly Native ancestry cohort). The prediction models were trained either on a highly Native (Blue) or highly European (Pink) cohort established from the CANDELA sample. For testing, we created equally-sized 4-folds for each pool of individuals. We built the RF models using three of the folds and evaluated prediction accuracy in the left-out fold from the Native ancestry cohort (see Supplementary figure S4).
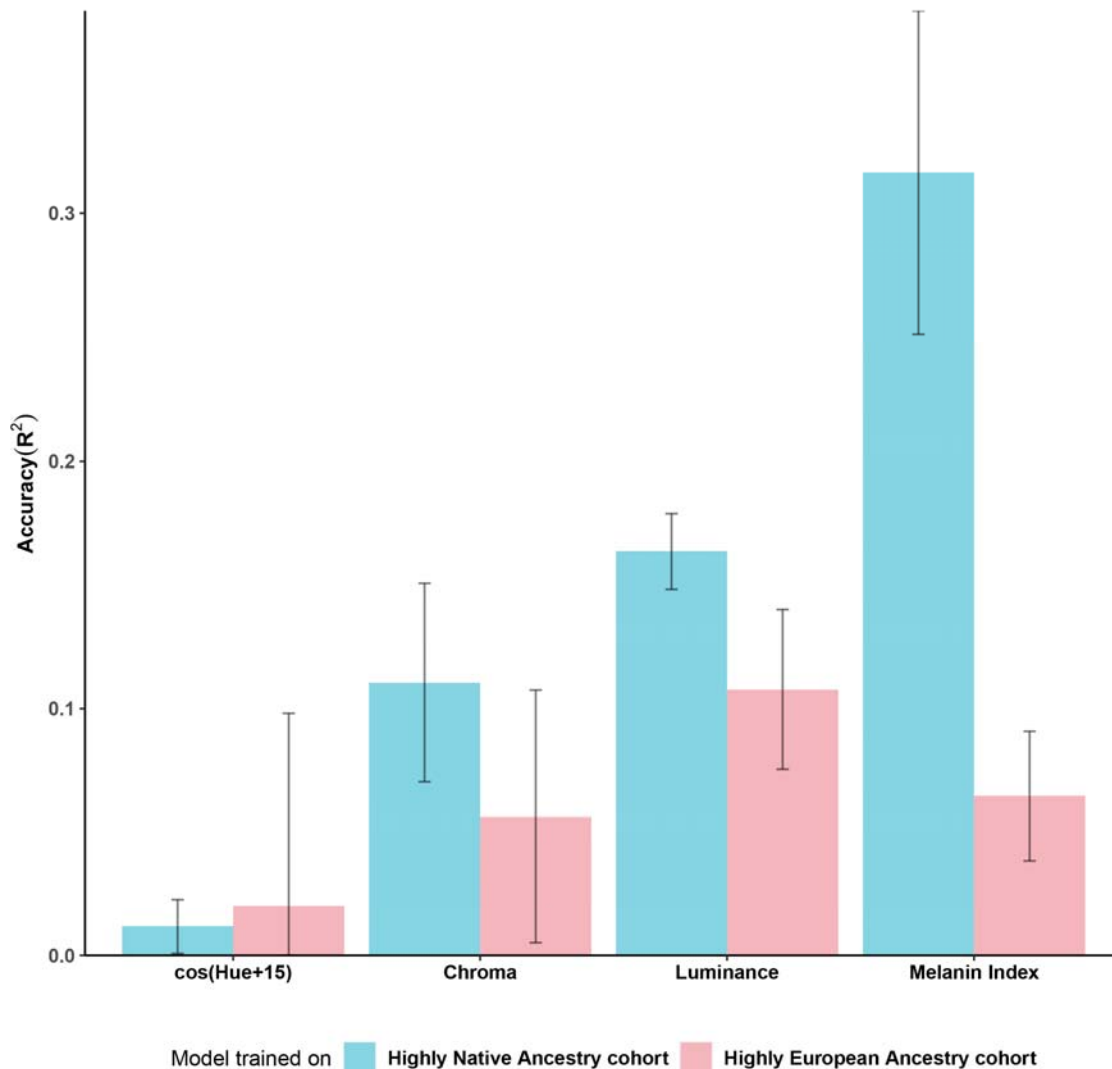

Prediction of skin pigmentation in Native Americans

We examined prediction performance of the CAN-S and HIrisPlex-S sets in a highly

Native dataset independent of CANDELA by predicting MI in 117 individuals from 17

Native American populations[40]. As above, we trained RF prediction models using

CANDELA individuals with >= 80% native ancestry. Since performance could not be measured directly in this dataset (due to the lack of phenotypic data), we examined the correlation of predicted skin pigmentation (MI) with solar radiation levels at the site of population sampling (Figure 6). Previous surveys of skin pigmentation in native populations from across the world have found a correlation between skin pigmentation and solar radiation[51], an observation that has been interpreted as the result of selection throughout human evolution. In the Native Americans examined here we obtained correlations of 0.543 (p-value 2 x 10^{-9}) and 0.163 (p-value 0.1) with the CAN-S and HIrisPlex-S SNP sets, respectively (Figure 6).

Figure 6. Solar radiation levels and skin pigmentation (MI) predicted with CAN-S and HIrisplex-S SNP sets. (a) Annual average of insolation incident on a horizontal surface (kWh/m^2/day - data from NASA Surface meteorology and Solar Energy, 2008) and location of the Native American population sampled. The predicted MI and solar radiation levels at the sampling site for 104 individuals from 16 Native American populations is shown in **(b)** CAN-S and **(c)** HIrisPlex-S.
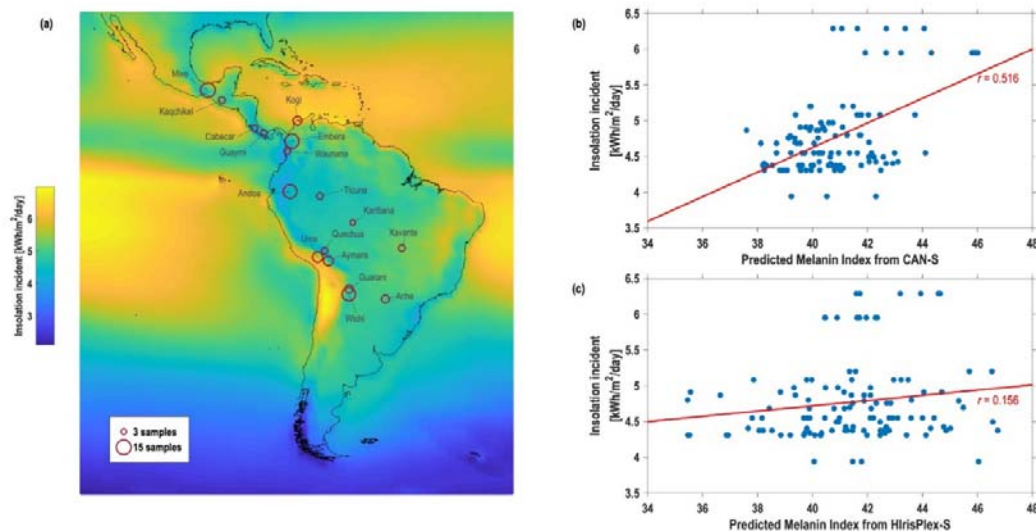
Discussion

Our results agree with previous studies in finding that regression and RF generally outperform other prediction methods (MLR being the approach implemented in HIrisPlex-S-Online). Owing to their tree-based structures, RF and XGB implicitly model the underlying genetic interaction across SNPs, and studies have noted the presence of epistasis (SNP-SNP interactions) between major pigmentation SNPs to some degree for various pigmentation traits, suggesting that prediction accuracy can be further increased by specifically allowing for interaction between SNPs[23,24,52]. Since tree-based models implicitly model epistasis including complex higher-order interactions, the difference of their prediction accuracy to additive linear/logistic models may also shed light on the genetic architecture of the traits. It is tempting to hypothesize that SNP interactions may have a more substantial impact on the genetic architecture of Hue and Chroma phenotypes[31,53], especially considering that these traits are relatively more non-linear in nature (e.g. Hue is an angle, i.e. a circular trait)[52]. Furthermore, the differences in prediction accuracy are much larger from linear to tree-based models (up to +0.19 gain in R^2 – see Supplementary table S3) than they are from HIrisPlex-S to CAN-E (up to +0.014 gain in R^2) whereas these two SNP sets only have 5 SNPs in common. We might therefore expect that such relevant interactions would be limited to a handful of SNPs, consistent with the literature which primarily observed interactions between major pigmentation SNPs[23,24,31]. Conversely, we may expect a greater additive polygenicity[23,25] in the case of skin pigmentation. Linear models overperform tree-based models for that trait (Figure 3) and the number of SNPs in the model has an impact on the prediction accuracy, as suggests the +0.07 gain in R^2 from HIrisPlex-S (34 SNPs) to CAN-S (120 SNPs). In line with these observations, we also noted that prediction accuracies are relatively lower for Melanin Index than for other continuous traits when constraining the number of SNPs in the model (see Supplementary table S6 and S7).

Previous studies[31,36,54,55] have shown that genetic ancestry correlates with pigmentation, probably resulting from the variable allele frequency across continents of certain pigmentation-related alleles. Consistent with this, we observe that continental genetic ancestry has significant predictive power (Figure 3, Supplementary table S3) and that the SNPs retained for prediction rank among the most correlated to continental ancestry components (see Supplementary table S2). Inclusion in the prediction models of pigmentation SNP sets selected from the GWAS further increases predictive power, markedly for quantitative pigmentation traits. Candidate SNPs on an average add twice the prediction accuracy of that brought in by the genetic ancestry for the quantitative traits. For the categorical traits, the difference is smaller but comparatively candidate SNPs decreases the proportion of incorrect classification by a larger amount than does continental genetic ancestry, especially for eye color.

The difference in predictive power for categorical traits is partly the result of their intrinsic lower informativeness relative to quantitative traits. This is particularly so for hair and eye color, as the CANDELA sample has a highly skewed phenotypic distribution for these traits: 82% of individuals in the CANDELA sample are assigned to the darkest category for eye and 85% for hair color. This skewed phenotypic distribution is consistent with the fact that world-wide lightly pigmented eyes and hair are essentially Western Eurasian traits, while there is considerable variation in skin pigmentation outside Europe[23–28]. That is, the occurrence of lightly pigmented hair and eyes in the CANDELA sample is essentially a reflection of its partly European ancestry. This may explain why the HIrisPlex-S SNP set built primarily in Europeans performs equally well to the CAN SNP set built on our more diverse admixed samples, and matches what has been observed in the literature for predicting eye and hair color in non-European populations[23,32,33].

By contrast, prediction of skin pigmentation is influenced by the genetic architecture of this trait in all three populations contributing to admixture in Latin America (Native Americans, Europeans and Africans). Consistent with this, our analysis along a Native American-European ancestry gradient show the highest gain in prediction accuracy for the ancestry bins with the greatest phenotypic diversity: that is, the highest European bin for eye/hair color and the highest Native American bin for skin color. In the bin with lowest European ancestry, there is hardly any gain in prediction accuracy for eye/hair color over the deterministic maxP strategy, as almost all individuals in that bin are in the highly pigmented category. On the other hand, this bin has the highest variation in skin pigmentation, also shows the greatest accuracy for the CAN-S, and this SNP set strongly outperforms HIrisPlex-S prediction for this ancestry bin (Figure 4) – all these facts together emphasize that skin color prediction in admixed Latin Americans is substantially influenced by the genetic architecture of this trait in non-European populations. Our observation of a stronger correlation of predicted skin with solar radiation levels in pigmentation in Native Americans for the CAN-S set, relative to HIrisPlex-S is also consistent with this, and with literature reporting comparatively poor portability of European-based pigmentation prediction models on non-European populations[23,32,33].

Although quantitative traits are intrinsically more informative than categorical ones, anthropological and forensic applications of pigmentation prediction are usually interested in a few discrete categories, often just two (e.g. blue v. non-blue eyes) or three (light, intermediate or dark color). In that setting we find that for the CANDELA dataset there is little difference in the prediction performance of HIrisPlex-S (which, in addition, was shrunken from 41 SNPs to 34 polymorphic SNPs in our dataset), relative to the CAN SNP sets developed here. However, we observe a strong impact of the training dataset used for optimizing the prediction methods, both in the extreme scenario explored in Figure 5, and

when comparing the prediction accuracies we obtained here with those published previously (Table 1). These observations question the portability across human populations of models developed using a training dataset with a different genetic makeup. This is most sharply illustrated by the difference in performance between the online implementation of HIrisPlex-S and the results we obtained when training models with this SNP set in the CANDELA data, particularly for skin pigmentation categories (Table 1). These observations thus caution against the use of the online implementation of HIrisPlex-S for the prediction of pigmentation traits in non-Europeans.

References

[1] T.E. King, G.G. Fortes, P. Balaresque, M.G. Thomas, D. Balding, P.M. Delser, R. Neumann, W. Parson, M. Knapp, S. Walsh, L. Tonasso, J. Holt, M. Kayser, J. Appleby, P. Forster, D. Ekserdjian, M. Hofreiter, K. Schürer, Identification of the remains of King Richard III, Nat Commun. 5 (2014) 5631. https://doi.org/10.1038/ncomms6631.

[2] I. Olalde, M.E. Allentoft, F. Sánchez-Quinto, G. Santpere, C.W.K. Chiang, M. DeGiorgio, J. Prado-Martinez, J.A. Rodríguez, S. Rasmussen, J. Quilez, O. Ramírez, U.M. Marigorta, M. Fernández-Callejo, M.E. Prada, J.M.V. Encinas, R. Nielsen, M.G. Netea, J. Novembre, R.A. Sturm, P. Sabeti, T. Marquès-Bonet, A. Navarro, E. Willerslev, C. Lalueza-Fox, Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European, Nature. 507 (2014) 225–228. https://doi.org/10.1038/nature12960.

[3] C. Lalueza-Fox, H. Rompler, D. Caramelli, C. Staubert, G. Catalano, D. Hughes, N. Rohland, E. Pilli, L. Longo, S. Condemi, M. de la Rasilla, J. Fortea, A. Rosas, M. Stoneking, T. Schoneberg, J. Bertranpetit, M. Hofreiter, A Melanocortin 1 Receptor Allele Suggests Varying Pigmentation Among Neanderthals, Science. 318 (2007) 1453–1455. https://doi.org/10.1126/science.1147417.

[4] M. Kayser, F. Liu, A.C.J.W. Janssens, F. Rivadeneira, O. Lao, K. van Duijn, M. Vermeulen, P. Arp, M.M. Jhamai, W.F.J. van IJcken, J.T. den Dunnen, S. Heath, D. Zelenika, D.D.G. Despriet, C.C.W. Klaver, J.R. Vingerling, P.T.V.M. de Jong, A. Hofman, Y.S. Aulchenko, A.G. Uitterlinden, B.A. Oostra, C.M. van Duijn, Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene, The American Journal of Human Genetics. 82 (2008) 411–423. https://doi.org/10.1016/j.ajhg.2007.10.003.

[5] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S.A. Gudjonsson, A. Palsson, G. Thorleifsson, S. Pálsson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, S.H. Vermeulen, A.M. Goldstein, M.A. Tucker, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Two newly identified genetic determinants of pigmentation in Europeans, Nat Genet. 40 (2008) 835–837. https://doi.org/10.1038/ng.160.

[6] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S.

Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans, Nat Genet. 39 (2007) 1443–1452. https://doi.org/10.1038/ng.2007.13.

[7] S. Walsh, F. Liu, K.N. Ballantyne, M. van Oven, O. Lao, M. Kayser, IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, Forensic Science International: Genetics. 5 (2011) 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004.

[8] S. Walsh, A. Wollstein, F. Liu, U. Chakravarthy, M. Rahu, J.H. Seland, G. Soubrane, L. Tomazzoli, F. Topouzis, J.R. Vingerling, J. Vioque, A.E. Fletcher, K.N. Ballantyne, M. Kayser, DNA-based eye colour prediction across Europe with the IrisPlex system, Forensic Science International: Genetics. 6 (2012) 330–340. https://doi.org/10.1016/j.fsigen.2011.07.009.

[9] J. Purps, M. Geppert, M. Nagy, L. Roewer, Evaluation of the IrisPlex eye colour prediction tool in a German population sample, Forensic Science International: Genetics Supplement Series. 3 (2011) e202–e203. https://doi.org/10.1016/j.fsigss.2011.08.101.

[10] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C.J.W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes, Current Biology. 19 (2009) R192–R193. https://doi.org/10.1016/j.cub.2009.01.027.

[11] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Science International: Genetics. 7 (2013) 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005.

[12] S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Science International: Genetics. 9 (2014) 150–161. https://doi.org/10.1016/j.fsigen.2013.12.006.

[13] S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pospiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser, Global skin colour prediction from DNA, Hum Genet. 136 (2017) 847–863. https://doi.org/10.1007/s00439-017-1808-5.

[14] L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pośpiech, M. Kukla-Bartoszek, T. Sijen, P. de Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh, The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, Forensic Science International: Genetics. 35 (2018) 123–135. https://doi.org/10.1016/j.fsigen.2018.04.004.

[15] K. Breslin, B. Wills, A. Ralf, M. Ventayol Garcia, M. Kukla-Bartoszek, E. Pospiech, A. Freire-Aradas, C. Xavier, S. Ingold, M. de La Puente, K.J. van der Gaag, N. Herrick, C. Haas, W. Parson, C. Phillips, T. Sijen, W. Branicki, S. Walsh, M. Kayser, HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms, Forensic Science International: Genetics. 43 (2019) 102152. https://doi.org/10.1016/j.fsigen.2019.102152.

[16] S.I. Candille, D.M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N.A. Garrison, J.Z. Li, R.M. Myers, G.S. Barsh, H. Tang, M.D. Shriver, Genome-Wide Association Studies of Quantitatively Measured Skin, Hair, and Eye Pigmentation in Four European Populations, PLoS ONE. 7 (2012) e48294. https://doi.org/10.1371/journal.pone.0048294.

[17] K. Zaorska, P. Zawierucha, M. Nowicki, Prediction of skin color, tanning and freckling from DNA in Polish population: linear regression, random forest and neural network

approaches, Hum Genet. 138 (2019) 635–647. https://doi.org/10.1007/s00439-019-02012-w.

[18] E. Pośpiech, Y. Chen, M. Kukla-Bartoszek, K. Breslin, A. Aliferi, J.D. Andersen, D. Ballard, L. Chaitanya, A. Freire-Aradas, K.J. van der Gaag, L. Girón-Santamaría, T.E. Gross, M. Gysi, G. Huber, A. Mosquera-Miguel, C. Muralidharan, M. Skowron, Á. Carracedo, C. Haas, N. Morling, W. Parson, C. Phillips, P.M. Schneider, T. Sijen, D. Syndercombe-Court, M. Vennemann, S. Wu, S. Xu, L. Jin, S. Wang, G. Zhu, N.G. Martin, S.E. Medland, W. Branicki, S. Walsh, F. Liu, M. Kayser, Towards broadening Forensic DNA Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA, Forensic Science International: Genetics. 37 (2018) 241–251. https://doi.org/10.1016/j.fsigen.2018.08.017.

[19] V. Kastelic, K. Drobnič, A single-nucleotide polymorphism (SNP) multiplex system: the association of five SNPs with human eye and hair color in the Slovenian population and comparison using a Bayesian network and logistic regression model, Croat Med J. 53 (2012) 401–408. https://doi.org/10.3325/cmj.2012.53.401.

[20] K.L. Hart, S.L. Kimura, V. Mushailov, Z.M. Budimlija, M. Prinz, E. Wurmbach, Improved eye- and skin-color prediction based on 8 SNPs, Croat Med J. 54 (2013) 248–256. https://doi.org/10.3325/cmj.2013.54.248.

[21] V. Kastelic, E. Pośpiech, J. Draus-Barini, W. Branicki, K. Drobnič, Prediction of eye color in the Slovenian population using the IrisPlex SNPs, Croat Med J. 54 (2013) 381–386. https://doi.org/10.3325/cmj.2013.54.381.

[22] P. Dario, H. Mouriño, A.R. Oliveira, I. Lucas, T. Ribeiro, M.J. Porto, J. Costa Santos, D. Dias, F. Corte Real, Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population, Int J Legal Med. 129 (2015) 1191–1200. https://doi.org/10.1007/s00414-015-1248-5.

[23] E. Pośpiech, A. Wojas-Pelc, S. Walsh, F. Liu, H. Maeda, T. Ishikawa, M. Skowron, M. Kayser, W. Branicki, The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction, Forensic Science International: Genetics. 11 (2014) 64–72. https://doi.org/10.1016/j.fsigen.2014.01.012.

[24] E.E. Quillen, H.L. Norton, E.J. Parra, F. Lona☐Durazo, K.C. Ang, F.M. Illiescu, L.N. Pearson, M.D. Shriver, T. Lasisi, O. Gokcumen, I. Starr, Y. Lin, A.R. Martin, N.G. Jablonski, Shades of complexity: New perspectives on the evolution and genetic architecture of human skin, Am J Phys Anthropol. 168 (2019) 4–26. https://doi.org/10.1002/ajpa.23737.

[25] A.R. Martin, M. Lin, J.M. Granka, J.W. Myrick, X. Liu, A. Sockell, E.G. Atkinson, C.J. Werely, M. Möller, M.S. Sandhu, D.M. Kingsley, E.G. Hoal, X. Liu, M.J. Daly, M.W. Feldman, C.R. Gignoux, C.D. Bustamante, B.M. Henn, An Unexpectedly Complex Architecture for Skin Pigmentation in Africans, Cell. 171 (2017) 1340-1353.e14. https://doi.org/10.1016/j.cell.2017.11.015.

[26] L. Deng, S. Xu, Adaptation of human skin color in various populations, Hereditas. 155 (2018) 1. https://doi.org/10.1186/s41065-017-0036-2.

[27] H.L. Norton, M. Edwards, S. Krithika, M. Johnson, E.A. Werren, E.J. Parra, Quantitative assessment of skin, hair, and iris variation in a diverse sample of individuals and associated genetic variation: QUANTITATIVE ASSESSMENT OF PIGMENTARY PHENOTYPE, Am. J. Phys. Anthropol. 160 (2016) 570–581. https://doi.org/10.1002/ajpa.22861.

[28] M. Jonnalagadda, M.A. Faizan, S. Ozarkar, R. Ashma, S. Kulkarni, H.L. Norton, E. Parra, A Genome-Wide Association Study of Skin and Iris Pigmentation among

Individuals of South Asian Ancestry, Genome Biology and Evolution. 11 (2019) 1066–1076. https://doi.org/10.1093/gbe/evz057.

[29] M. Edwards, D. Cha, S. Krithika, M. Johnson, G. Cook, E.J. Parra, Iris pigmentation as a quantitative trait: variation in populations of European, East Asian and South Asian ancestry and association with candidate gene polymorphisms, Pigment Cell Melanoma Res. 29 (2016) 141–162. https://doi.org/10.1111/pcmr.12435.

[30] L. Rawofi, M. Edwards, S. Krithika, P. Le, D. Cha, Z. Yang, Y. Ma, J. Wang, B. Su, L. Jin, H.L. Norton, E.J. Parra, Genome-wide association study of pigmentary traits (skin and iris color) in individuals of East Asian ancestry, PeerJ. 5 (2017) e3951. https://doi.org/10.7717/peerj.3951.

[31] K. Adhikari, J. Mendoza-Revilla, A. Sohail, M. Fuentes-Guajardo, J. Lampert, J.C. Chacón-Duque, M. Hurtado, V. Villegas, V. Granja, V. Acuña-Alonzo, C. Jaramillo, W. Arias, R.B. Lozano, P. Everardo, J. Gómez-Valdés, H. Villamil-Ramírez, C.C. Silva de Cerqueira, T. Hunemeier, V. Ramallo, L. Schuler-Faccini, F.M. Salzano, R. Gonzalez-José, M.-C. Bortolini, S. Canizales-Quinteros, C. Gallo, G. Poletti, G. Bedoya, F. Rothhammer, D.J. Tobin, M. Fumagalli, D. Balding, A. Ruiz-Linares, A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia, Nat Commun. 10 (2019) 358. https://doi.org/10.1038/s41467-018-08147-0.

[32] A. Pneuman, Z.M. Budimlija, T. Caragine, M. Prinz, E. Wurmbach, Verification of eye and skin color predictors in various populations, Legal Medicine. 14 (2012) 78–83. https://doi.org/10.1016/j.legalmed.2011.12.005.

[33] L. Yun, Y. Gu, H. Rajeevan, K.K. Kidd, Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations, Int J Legal Med. 128 (2014) 447–453. https://doi.org/10.1007/s00414-013-0953-1.

[34] H.L. Norton, R.A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V.A. Canfield, D.G. Bradley, B. McEvoy, M.D. Shriver, Genetic Evidence for the Convergent Evolution of Light Skin in Europeans and East Asians, Molecular Biology and Evolution. 24 (2006) 710–722. https://doi.org/10.1093/molbev/msl203.

[35] M. Edwards, A. Bigham, J. Tan, S. Li, A. Gozdzik, K. Ross, L. Jin, E.J. Parra, Association of the OCA2 Polymorphism His615Arg with Melanin Content in East Asian Populations: Further Evidence of Convergent Evolution of Skin Pigmentation, PLoS Genet. 6 (2010) e1000867. https://doi.org/10.1371/journal.pgen.1000867.

[36] A. Ruiz-Linares, K. Adhikari, V. Acuña-Alonzo, M. Quinto-Sanchez, C. Jaramillo, W. Arias, M. Fuentes, M. Pizarro, P. Everardo, F. de Avila, J. Gómez-Valdés, P. León-Mimila, T. Hunemeier, V. Ramallo, C.C. Silva de Cerqueira, M.-W. Burley, E. Konca, M.Z. de Oliveira, M.R. Veronez, M. Rubio-Codina, O. Attanasio, S. Gibbon, N. Ray, C. Gallo, G. Poletti, J. Rosique, L. Schuler-Faccini, F.M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, F. Rothhammer, G. Bedoya, D. Balding, R. Gonzalez-José, Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals, PLoS Genet. 10 (2014) e1004572. https://doi.org/10.1371/journal.pgen.1004572.

[37] K. Adhikari, J. Mendoza-Revilla, J.C. Chacón-Duque, M. Fuentes-Guajardo, A. Ruiz-Linares, Admixture in Latin America, Current Opinion in Genetics & Development. 41 (2016) 106–114. https://doi.org/10.1016/j.gde.2016.09.003.

[38] K. Adhikari, M. Fuentes-Guajardo, M. Quinto-Sánchez, J. Mendoza-Revilla, J. Camilo Chacón-Duque, V. Acuña-Alonzo, C. Jaramillo, W. Arias, R.B. Lozano, G.M. Pérez, J. Gómez-Valdés, H. Villamil-Ramírez, T. Hunemeier, V. Ramallo, C.C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F.M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, M. Cheeseman, J. Rosique, G. Bedoya, F. Rothhammer, D. Headon, R. González-José, D. Balding, A. Ruiz-Linares, A

genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation, Nat Commun. 7 (2016) 11616. https://doi.org/10.1038/ncomms11616.

[39] K. Adhikari, T. Fontanil, S. Cal, J. Mendoza-Revilla, M. Fuentes-Guajardo, J.-C. Chacón-Duque, F. Al-Saadi, J.A. Johansson, M. Quinto-Sanchez, V. Acuña-Alonzo, C. Jaramillo, W. Arias, R. Barquera Lozano, G. Macín Pérez, J. Gómez-Valdés, H. Villamil-Ramírez, T. Hunemeier, V. Ramallo, C.C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F.M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, F. Rothhammer, G. Bedoya, R. Gonzalez-José, D. Headon, C. López-Otín, D.J. Tobin, D. Balding, A. Ruiz-Linares, A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features, Nat Commun. 7 (2016) 10815. https://doi.org/10.1038/ncomms10815.

[40] J.-C. Chacón-Duque, K. Adhikari, M. Fuentes-Guajardo, J. Mendoza-Revilla, V. Acuña-Alonzo, R. Barquera, M. Quinto-Sánchez, J. Gómez-Valdés, P. Everardo Martínez, H. Villamil-Ramírez, T. Hünemeier, V. Ramallo, C.C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, M. Villena, R. Vásquez, E. Llop, J.R. Sandoval, A.A. Salazar-Granara, M.-L. Parolin, K. Sandoval, R.I. Peñaloza-Espinosa, H. Rangel-Villalobos, C.A. Winkler, W. Klitz, C. Bravi, J. Molina, D. Corach, R. Barrantes, V. Gomes, C. Resende, L. Gusmão, A. Amorim, Y. Xue, J.-M. Dugoujon, P. Moral, R. González-José, L. Schuler-Faccini, F.M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, G. Poletti, C. Gallo, G. Bedoya, F. Rothhammer, D. Balding, G. Hellenthal, A. Ruiz-Linares, Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance, Nat Commun. 9 (2018) 5388. https://doi.org/10.1038/s41467-018-07748-z.

[41] K. Adhikari, G. Reales, A.J.P. Smith, E. Konka, J. Palmen, M. Quinto-Sanchez, V. Acuña-Alonzo, C. Jaramillo, W. Arias, M. Fuentes, M. Pizarro, R. Barquera Lozano, G. Macín Pérez, J. Gómez-Valdés, H. Villamil-Ramírez, T. Hunemeier, V. Ramallo, C.C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F.M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, F. Rothhammer, G. Bedoya, R. Calderón, J. Rosique, M. Cheeseman, M.F. Bhutta, S.E. Humphries, R. Gonzalez-José, D. Headon, D. Balding, A. Ruiz-Linares, A genome-wide association study identifies multiple loci for variation in human ear morphology, Nat Commun. 6 (2015) 7500. https://doi.org/10.1038/ncomms8500.

[42] O. Maroñas, C. Phillips, J. Söchtig, A. Gomez-Tato, R. Cruz, J. Alvarez-Dios, M.C. de Cal, Y. Ruiz, M. Fondevila, Á. Carracedo, M.V. Lareu, Development of a forensic skin colour predictive test, Forensic Science International: Genetics. 13 (2014) 34–44. https://doi.org/10.1016/j.fsigen.2014.06.017.

[43] A. Freire-Aradas, Y. Ruiz, C. Phillips, O. Maroñas, J. Söchtig, A.G. Tato, J.Á. Dios, M.C. de Cal, V.N. Silbiger, A.D. Luchessi, A.D. Luchessi, M.A. Chiurillo, Á. Carracedo, M.V. Lareu, Exploring iris colour prediction and ancestry inference in admixed populations of South America, Forensic Science International: Genetics. 13 (2014) 3–9. https://doi.org/10.1016/j.fsigen.2014.06.007.

[44] W. Branicki, F. Liu, K. van Duijn, J. Draus-Barini, E. Pośpiech, S. Walsh, T. Kupiec, A. Wojas-Pelc, M. Kayser, Model-based prediction of human hair color using DNA variants, Hum Genet. 129 (2011) 443–454. https://doi.org/10.1007/s00439-010-0939-8.

[45] T.K.M. Leite, R.M.C. Fonseca, N.M. de França, E.J. Parra, R.W. Pereira, Genomic Ancestry, Self-Reported "Color" and Quantitative Measures of Skin Pigmentation in Brazilian Admixed Siblings, PLoS ONE. 6 (2011) e27162. https://doi.org/10.1371/journal.pone.0027162.

[46] A. Lo, H. Chernoff, T. Zheng, S.-H. Lo, Why significant variables aren't automatically good predictors, Proc Natl Acad Sci USA. 112 (2015) 13892–13897. https://doi.org/10.1073/pnas.1518285112.

[47] R.K. Valenzuela, M.S. Henderson, M.H. Walsh, N.A. Garrison, J.T. Kelch, O. Cohen-Barak, D.T. Erickson, F. John Meaney, J. Bruce Walsh, K.C. Cheng, S. Ito, K. Wakamatsu, T. Frudakis, M. Thomas, M.H. Brilliant, Predicting Phenotype from Genotype: Normal Pigmentation, Journal of Forensic Sciences. 55 (2010) 315–322. https://doi.org/10.1111/j.1556-4029.2009.01317.x.

[48] O. Spichenok, Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, E. Wurmbach, Prediction of eye and skin color in diverse populations using seven SNPs, Forensic Science International: Genetics. 5 (2011) 472–478. https://doi.org/10.1016/j.fsigen.2010.10.005.

[49] B. Alipanahi, P. Fontanillas, 23andMe Resarch Team, S. Pitts, R. Gentleman, Pigmentor—Accurate prediction of multiple pigmentation phenotypes, (2017).

[50] Y. Ruiz, C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, M. Casares de Cal, R. Cruz, O. Maroñas, J. Söchtig, M. Fondevila, M.J. Rodriguez-Cid, Á. Carracedo, M.V. Lareu, Further development of forensic eye color predictive tests, Forensic Science International: Genetics. 7 (2013) 28–40. https://doi.org/10.1016/j.fsigen.2012.05.009.

[51] N.G. Jablonski, G. Chaplin, The evolution of human skin coloration, Journal of Human Evolution. 39 (2000) 57–106. https://doi.org/10.1006/jhev.2000.0403.

[52] E. Pośpiech, J. Draus-Barini, T. Kupiec, A. Wojas-Pelc, W. Branicki, Gene–gene interactions contribute to eye colour variation in humans, J Hum Genet. 56 (2011) 447–455. https://doi.org/10.1038/jhg.2011.38.

[53] A. Wollstein, S. Walsh, F. Liu, U. Chakravarthy, M. Rahu, J.H. Seland, G. Soubrane, L. Tomazzoli, F. Topouzis, J.R. Vingerling, J. Vioque, S. Böhringer, A.E. Fletcher, M. Kayser, Novel quantitative pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye colour, Sci Rep. 7 (2017) 43359. https://doi.org/10.1038/srep43359.

[54] T.N. Frudakis, Molecular photofitting: predicting ancestry and phenotype using DNA, Elsevier/Academic Press, Amsterdam ; Boston, 2008.

[55] C.C.S. de Cerqueira, T. Hünemeier, J. Gomez-Valdés, V. Ramallo, C.D. Volasko-Krause, A.A.L. Barbosa, P. Vargas-Pinilla, R.C. Dornelles, D. Longo, F. Rothhammer, G. Bedoya, S. Canizales-Quinteros, V. Acuña-Alonzo, C. Gallo, G. Poletti, R. González-José, F.M. Salzano, S.M. Callegari-Jacques, L. Schuler-Faccini, A. Ruiz-Linares, M. Cátira Bortolini, for CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin America), Implications of the Admixture Process in Skin Color Molecular Assessment, PLoS ONE. 9 (2014) e96886. https://doi.org/10.1371/journal.pone.0096886.

Acknowledgments

Competing interests: The authors declare that they have no competing interests.

Data availability:. Summary statistics from the GWAS analyses on which were established

the three CAN SNP sets used in this study were previously deposited at GWAS central

(http://www.gwascentral.org/study/HGVST3308).

Tables

Table 1. Overall Accuracy (%) and trait AUC (%) for categorical eye (A), hair (B) and skin (C) color obtained here and in other studies (% frequency of trait in sample is shown in parentheses).

(A)

Sample	CANDELA N=6,529				Venezuela/Brazil[b] N = 99		Europeans[d] N=2,364	
SNP set		CAN-E[a]	HIrisPlex-S	HIrisPlexS-Online				
Overall Accuracy		89	89	87				83
1. Blue/gray	(3)	93	93	93	(12)	85	(68)	91
2. Intermediate	(13)	89	89	85	(23)	NA	(10)	73
3. Brown/black	(85)	92	92	90	(64)	91	(23)	93

(B)

Sample	CANDELA N=6,495				Europeans[c] N = 385	
SNP set		CAN-H	HIrisPlex-S	HIrisPlexS-Online		
Overall Accuracy		85	84	56		71
1.Red	(0)	NA	NA	NA	(25)	90
2. Blond	(3)	94	92	90	(54)	75
3. Brown	(16)	82	81	65	(9)	72
4. Black	(82)	87	85	80	(12)	78

(C)

Sample	CANDELA N=6,526				World-wide N=1,159[e]	
SNP set		CAN-S	HIrisPlex-S	HIrisPlex-S-Online		
Overall Accuracy		73	72	26		87
1. Fair	(38)	83	81	78	(92)	97
2. Intermediate	(59)	77	77	32	(3)	83
3. Dark	(3)	86	84	81	(5)	96

HIrisPlexS-Online: https://hirisplex.erasmusmc.nl

[a] Obtained using a model with only SNPs as predictors (no covariates) and with MLR as prediction method.

[b] Obtained using HIrisPlex as described in Freire-Aradas et al. 2014[43].

[c] Obtained using HIrisPlex-S as described in Branicki et al 2011[44], here accuracy is calculated from a 4-way classification setup.

[d] Obtained from the specificity reported in Liu et al 2009[10]

[e] This prediction values reported in Walsh et al 2017[13] were obtained from the prediction of the 194 samples from Maronas et al 2014[42]